Constraining the Inter-galactic Medium with the SDSS Lyα Forest

Khee-Gan (K.G.) Lee

Princeton University

November 30, 2010

Outline

Based on Lee & Spergel 2010 (arxiv:1007.3734)

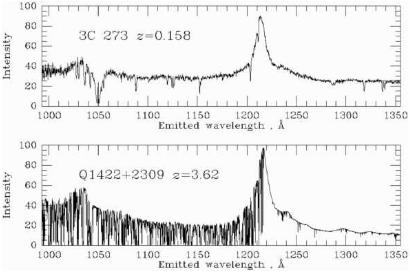
- The inter-galactic medium (IGM) comprises $> 80\%$ of baryons at $z > 1$
- The hydrogen Lyα forest is the main source of data on the IGM
- Most observational constraints on the IGM have been from 8m-class telescopes
  — Small numbers, $\sim 10 - 20$, high-resolution ($R \sim 30000$)
- SDSS represents $\sim 10^4$ Lyα forest sightlines at moderate resolution ($R = 2000$)

Can the SDSS Lyα forest contribute to understanding the IGM, and what statistics are useful for this?

Acknowledgements: David Spergel (Princeton), Sal Torquato (Chemistry/Applied Physics, Princeton), Matt McQuinn (UC Berkeley)
Special thanks to the Princeton Center for Theoretical Science

The Lyα Forest

Absorption of observed quasar flux at $\lambda_{rest} < 1216$Å by intervening neutral hydrogen along the line-of-sight, tracing underlying density fluctuations



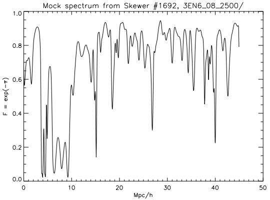
At $z \sim 3$, each line-of-sight probes $\sim 500\,\mathrm{h}^{-1}$ Mpc!

Important probe of the universe at $z > 1$

Basic Astrophysics of the Lyα Forest

We measure absorbed flux $F = e^{-\tau}$.
Assuming ionization equilibrium, can write down the fluctuating Gunn-Peterson approximation (FGPA):

$$\tau(x) \propto \frac{\bar{T}^{-0.7}}{\Gamma} \Delta(x)^{2-0.7(\gamma-1)}$$



Cosmology/LSS

- Matter density field, $\Delta(x) = \rho(x)/\bar{\rho}$
- Peculiar velocities, $v \leftrightarrow x$

IGM physics

- Photoionization rate, Γ
- Gas temperature, $\bar{T}$
- Gas equation of state, γ

Why should anyone care about the IGM?

IGM affects Lyα forest at $\sim$% level
— needs to be well-contrained to use Lyα forest for precision cosmology (matter power spectrum, neutrino masses etc)

The IGM astrophysics depends on (in no particular order):

- Star formation at $z \sim 3$
- Quasar physics – bias, luminosity function, duty cycle
- Gas clumping near sources $\rightarrow$ radiation filtering

And these are affected by IGM...

- Star formation/galaxy evolution
- Missing baryon problem at low-z

HeII reionization at $z \sim 3$ by quasars ?

Astrophysics of the IGM at $z \sim 3$

Post-reionization **fiducial** values from Hui & Gnedin 1997:

Temperature field, $\bar{T}$ (**Fiducial**: $\bar{T} \sim 10^4$)

- Temperature bears imprint of energy injection processes in the past
- Various reionization events: hydrogen ($z > 6$), HeII ($z \sim 3$)
- *Constraints: Lyα absorption profile fitting (Schaye et al. 2000, Becker et al. 2010)*

Photoionization rate, Γ (**Fiducial**: $\Gamma \sim 10^{-12} \mathrm{s}^{-1}$)

size

- The IGM is **highly ionized** from UV background radiation
- The UV background $\rightarrow$ filtered radiation from radiation sources in the universe
- *Constraints: Evolution of mean Lyα optical depth (Bernardi et al. 2003, Faucher-Giguère et al. 2008, Dall'Aglio et al. 2009)*

Astrophysics of the IGM at $z \sim 3$

Post-reionization **fiducial** values from Hui & Gnedin 1997:

Temperature field, $\bar{T}$ (**Fiducial**: $\bar{T} \sim 10^4$)

- Temperature bears imprint of energy injection processes in the past
- Various reionization events: hydrogen ($z > 6$), HeII ($z \sim 3$)
- *Constraints: Lyα absorption profile fitting (Schaye et al. 2000, Becker et al. 2010)*

Photoionization rate, Γ (**Fiducial**: $\Gamma \sim 10^{-12} s^{-1}$)

size

- The IGM is **highly ionized** from UV background radiation
- The UV background $\rightarrow$ filtered radiation from radiation sources in the universe
- *Constraints: Evolution of mean Lyα optical depth (Bernardi et al. 2003, Faucher-Giguère et al. 2008, Dall'Aglio et al. 2009)*

Astrophysics of the IGM at $z \sim 3$ (continued)

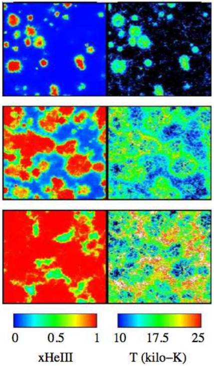
Temperature-density relation, γ (**Fiducial**: $\gamma = 1.6$)

- $T(\Delta) \propto \Delta^{\gamma-1}$
- Set by adiabatic heating/cooling, photoheating, Compton cooling
- *Constraints: Flux probability distribution function (McDonald et al. 2000, Becker et al. 2007, Bolton et al. 2008)*

See also HeII Gunn-Peterson trough studies from HST COS
— HeII reionization at $z \approx 2.7$? (Schull et al 2010)

Thermal Inhomogeneities

HeII reionization simulations from McQuinn et al. 2009:



429 h^{-1} Mpc box, 8.4 h^{-1} Mpc slices

Depends on QSO duty cycle + luminosity functions, gas clumping, etc. See also Gleser et al. 2005, Furlanetto & Oh 2008.

How can SDSS contribute to understanding the IGM?

Sloan Digital Sky Survey I/II

- $\sim$ 15000 quasars at $z > 2.2$
  — no worries about cosmic variance!
- Moderate resolution, R $\sim$ 2000
- Cannot use high-res techniques (Voight-profile fitting, wavelet decomposition)

Modelling steps

- Use 'off-the-shelf' Lyα forest sims
- Apply SDSS noise + systematics + numbers

Test out a few statistics to constrain IGM?

- Flux PDF
- Threshold probability functions

Mock Spectra

Used Martin White's $z = 2.5$ Franklin simulations (Slosar et al 2009) as basis
(`http://mwhite.berkeley.edu/BOSS/LyA/Franklin/`)

- DM particle mesh simulation $1500^3 \mathrm{Mpc}^3 \mathrm{h}^{-3}$, 3000^3 particles, 3000^3 cells at $z = 2.5$
- 150^2 Lyα skewers generated using FGPA with $\mathrm{T}_0 = 20000\mathrm{K}$, $\gamma - 1 = 0.5$
- Chop up each skewer to get ~rough length ($500\,\mathrm{h}^{-1}$ Mpc) of individual observed Lyα spectrum

Use FGPA to modify optical depths to make various toy models:

$$\tau \propto \bar{\mathrm{T}}^{-0.7} \Delta^{2-0.7(\gamma-1)} \left(1 + \frac{1}{\mathrm{H}(z)} \frac{dv_{\mathrm{pec}}}{dx}\right)$$

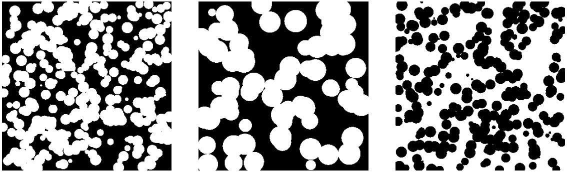
Toy Models

Generated homogeneous models with different γ:

- G1.5: $\gamma = 1.5$
- G1.3: $\gamma = 1.3$
- G0.8: $\gamma = 0.8$ (Inverted eq. of state)

Toy model for HeII inhomogeneities at z=2.5 (suggested by M. McQuinn):

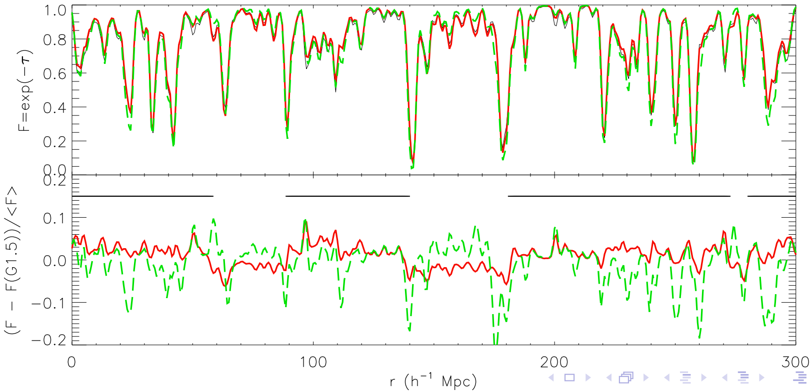
- Randomly throw down spheres into box at 50% volume fraction
- Inside sphere: $T_0 = 25000, \gamma - 1 = 0.2$
- Outside sphere: $T_0 = 15000, \gamma - 1 = 0.5$
- Left to right: R=50 (Model R50), R100, I50



Spot the difference!

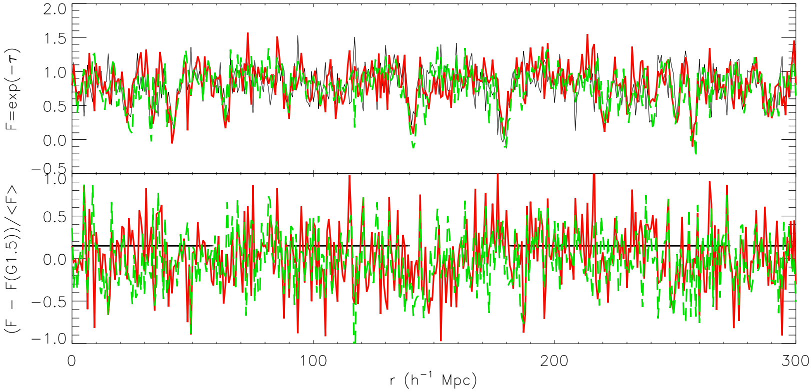
- Black: Homogeneous model G1.5, $\gamma = 1.5$
- Green: Homogeneous model G0.8, $\gamma = 0.8$
- Red: Inhomogeneous model R50, $50\,\mathrm{h}^{-1}$ Mpc hot bubbles

No Noise included in this plot!



Spot the difference! Part II

With S/N = 4 per pixel:



Should I even bother?

Anonymous English astronomer to me:

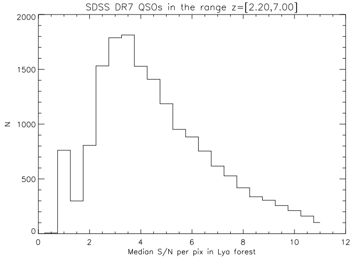
Don't waste your time!

SDSS Mock Sample

Sloan Digital Sky Survey DR7

Final Data Release 7 published in 2008.

- Schneider et al 2010 QSO catalog $\rightarrow$ 105,783 objects
- 17,582 quasars with $z > 2.2$



Mock sample:

- Apply redshift cut ($2.4 \geqslant z \geqslant 2.7$) and S/N cut (S/N > 4 per pixel) $\rightarrow$ 1500 quasars
- Each mock spectrum degraded to $R = \lambda/\Delta\lambda = 2000$ and given noise corresponding to S/N $= 4$ per pixel
- Errors in continuum fitting parametrized by allowing Gaussian spread σ_F in mean flux of individual spectra

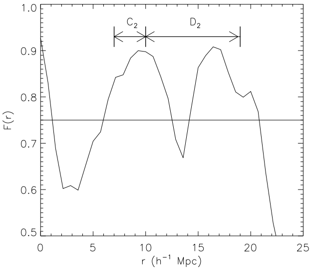
A new statistic for astronomy

Threshold Probability Function: (Torquato et al. J. Chem. Phys. 1988)

$$S_2(r) = C_2(r) + D_2(r)$$

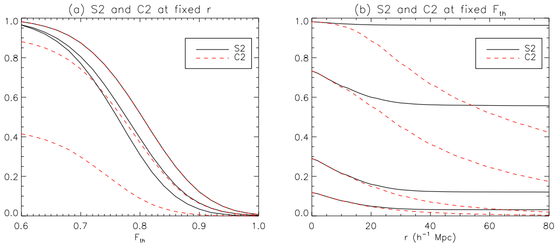
Threshold Probability Functions

- $S_2 \equiv$ prob. of finding 2 points above threshold at distance r
- $C_2 \equiv$ prob. of finding 2 points above threshold **in the same cluster** at distance r
- $D_2 \equiv$ prob. of finding 2 point above threshold **in other clusters** at distance r



For Lyα spectra, define phases through flux thresholds F_{th}
$\rightarrow$ 2-dimensional function: $S_2 \equiv S_2(r, F_{th})$

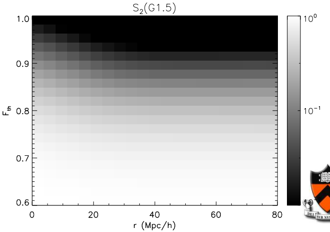
S_2 evaluated on DR7 mock sample (smoothed on $\sigma = 10\,h^{-1}\,\mathrm{Mpc}$)



Some identities:

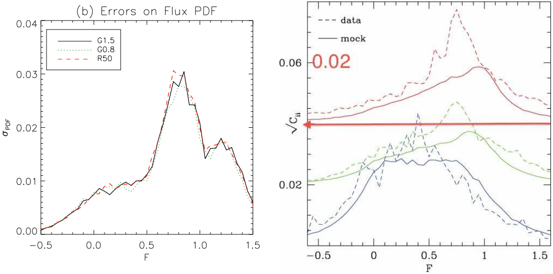
- $S_2(r = 0|F_{th})$ is the integral of flux PDF, $S_2(0|F_{th}) = \int_{F_{th}}^{1} p(F)dF$
- In the absence of long-range order $S_2(r|F_{th}) \to S_2^2(0|F_{th})$

Think of $S_2(r, F_{th})$ as the **flux PDF with spatial information included**



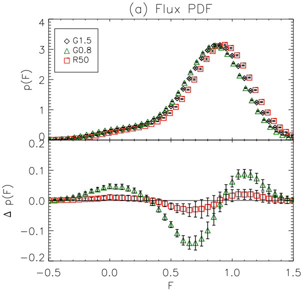
Systematics and Errors

- TPF is related to flux PDF, so compute flux PDF as a check on errors
- Flux continuum errors: 10% Gaussian spread in normalization of mock spectra
- Compare errors on flux PDF (left) with Desjacques et al. 07 (right, used DR3 data) to check errors



Constraining γ with flux PDF

- Flux PDF from SDSS constrain γ
- Computed
  $-\ln\mathcal{L} = \frac{1}{2}(x_i-\mu_i)^T C_{ij}^{-1}(x_j-\mu_j)$
  ~reduced Chi-squared, but want to *maximize* values
- Inhomogeneous models degenerate with homogeneous IGM with $\gamma = 1.3$
- But otherwise get good handle on γ if assume homogeneous IGM...



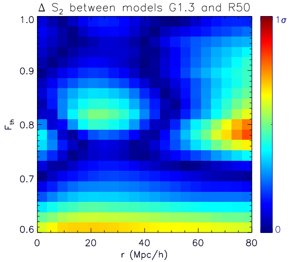
	Homogeneous models			Inhomogeneous models		
	G1.5	G1.3	G0.8	R50	R100	I50
G1.5	0.0	21.8	388.5	15.4	13.1	18.1
G1.3	...	0.0	154.1	2.3	0.3	2.3
G0.8	...	...	0.0	209.1	161.6	218.7
R50	...	...	...	0.0	0.4	0.2
R100	...	...	...	...	...	0.2
I50	...	...	...	...	...	0.0

Constraints with SDSS flux PDF

- Flux PDF from DR7 can distinguish between $\gamma = 1.5$ and $\gamma = 1.3$ with $-\ln \mathcal{L} \approx 20$.
- Viel et al. 2009 reported $\gamma = 0.7 \pm 0.2$ at $z = 3$ from flux PDF of high-res spectra.
- Likelihoods previously shown are for 1500 S/N = 4 skewers
  - These are the numbers of spectra in DR7 at $2.4 \geqslant z \geqslant 2.7$
  - In DR7 have $\gtrsim 700$ quasars in $\Delta z = 0.3$ bins up to $z \sim 3.5$
  - $-\ln \mathcal{L}$ is roughly proportional to sample size, so get $\Delta\gamma \approx 0.2$ with $-\ln \mathcal{L} \approx 10$ across these z-bins

$\rightarrow$ **Can measure evolution of γ across the epoch of HeII reionization**

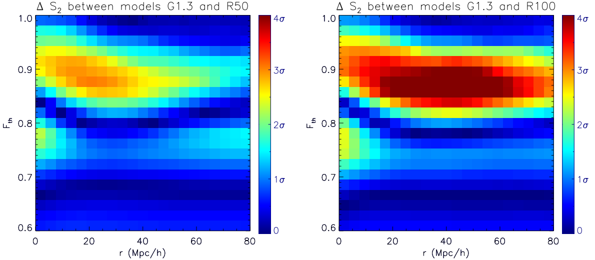
Threshold Probability Functions for Different IGM Models



$-\ln\mathcal{L}$ for $S_2(r, F_{th})$, assuming DR7 sample at $z = 2.5$, **10% flux errors**

	Homogeneous models			Inhomogeneous models		
	G1.5	G1.3	G0.8	R50	R100	I50
G1.5	0.0	54.1	837.6	67.8	37.6	69.1
G1.3	...	0.0	400.6	14.1	13.9	17.7
G0.8	...	...	0.0	439.1	352.0	389.2
R50	...	...	...	0.0	6.1	3.2
R100	...	...	...	...	0.0	14.8
I50	...	...	...	...	...	0.0

If we can get 3% accuracy in flux continuum...



— $\ln \mathcal{L}$ for $S_2(r, F_{th})$ assuming DR7 sample at $z = 2.5$, **3% flux errors**

	Homogeneous models			Inhomogeneous models		
	G1.5	G1.3	G0.8	R50	R100	I50
G1.5	0.0	150.1	1620.7	205.3	292.7	191.1
G1.3	...	0.0	779.6	29.0	65.8	28.6
G0.8	...	...	0.0	756.8	1632.1	965.9
R50	...	...	...	0.0	16.2	2.7
R100	...	...	...	...	0.0	29.8
I50	...	...	...	...	...	0.0

Summary/Conclusions

I introduced the threshold probability function statistics, S_2, C_2, and D_2, that is basically the flux PDF with spatial information

- Tested them with mock spectra assuming simple toy models for the IGM and SDSS DR7 data
- With 10% flux continuum error, can make detection of temperature inhomogeneities
- With 3% flux continuum error, can constrain the scale of temperature inhomogeneities
- Also found that the flux PDF from SDSS can make interesting constraints on homogeneous γ across the epoch of HeII reioization

See also Lee & Spergel 2010, arxiv:1007.3734

Next: Figure out how to deal with systematics (continuum + metals) then measure PDF + TPF from SDSS DR7

Next Steps: Dealing with systematics

Metal line contamination

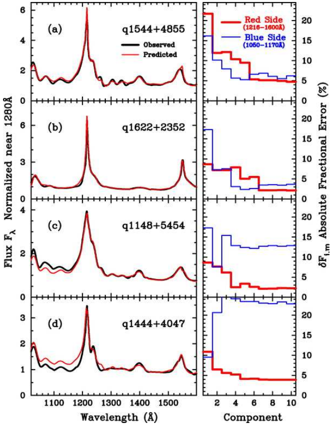
- Use Renyue Cen's hydro simulatons (Cen & Chisari 2010)
- Tracks redshifts down to $z = 0$
- Includes metals from SF feedback.

Continuum fitting

- Use PCA continuum fitting with mean flux regulation
- Few % continuum fitting errors, close to level of dispersion in $\bar{F}$ across lines-of-sight
- In collaboration with Nao Suzuki (LBNL), Shirley Ho (LBNL) and Brice Menard (JHU)

PCA Continuum Fitting

- Continuum fitting (along with metal line contamination) are the greatest systematics in Lyα forest research
- For moderate-resolution data, need to extrapolate the continuum from redwards of the quasar Lyα emission line
- Suzuki et al. 2005 established PCA techniques for this by using 50 low-z HST spectra
- Shape of the continuum is well predicted, but $\sim 10\% errors$ in the normalization



Suzuki et al. 2005

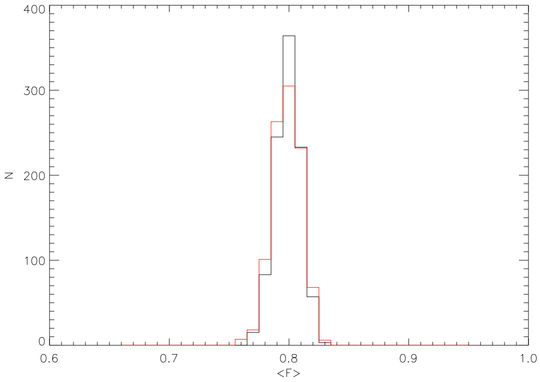
Mean-Flux Regulation

Suzuki et al. used **only** redwards of Lyα in their continuum predictions, but we do have some info in the Lyα forest region itself

1. Do PCA prediction based on red side of spectrum i
2. Extract Lyα forest based on this estimate, then calculate mean transmission $\overline{F}_i$ for the spectrum.
3. Compare with the **global** mean flux $\langle F(z) \rangle$
4. Demand that $\overline{F}_i = \langle F(z) \rangle$, and recalculate continuum based on this

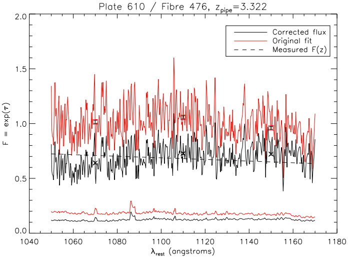
Dispersion between lines-of-sight

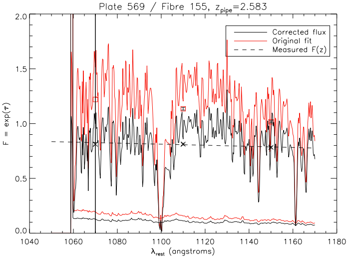
Mock skewers generated from Martin White's Franklin sims (same as Lee & Spergel), global $< F >$ set to 0.8



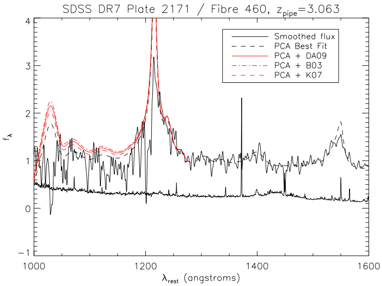
Application to DR7 — Mean-flux Regulation

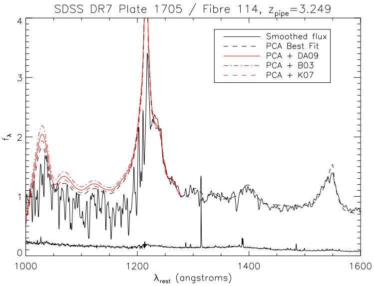
1. Take PCA fit from previous steps, extract Lyα forest ($1050\text{Å} < \lambda < 1170\text{Å}$)
2. Compute mean flux in 3 bins within the forest
3. Introduce an new exponential component to the model spectrum: $f_{exp} = \exp\left(-\delta_{exp}(\lambda - 1280\text{Å})/1280\text{Å}\right)$
4. Fit exponential such that mean flux bins match published $< F > (z)$ measurements





Mean-flux Regulation — Results





Outstanding issue: $\sim 5\%$ spread in values of $< F > (z)$ reported in the literature